Extracting the Critical Frequency Bands to Classify Vigilance States of Rats by Using a Novel Feature Selection Algorithm

Chien-Hsing Chou, Chung-Chih Kuo, Zong-En Yu, Hsien-Pang Tai1, Ke-Wei Chen

***Abstract*—Identifying mammalian vigilance states has recently become an important topic in biological science research. The biological researchers concern not only to improve the accuracy rate for classifying the vigilance states, but also to extract the meaningful frequency bands. In this study, we propose a novel feature selection to extract the critical frequency bands of rat's EEG signals. The proposed algorithm adopts the concept of neighborhood relation during adding and eliminating a candidate feature. In the experiments, the proposed method shows better accuracy rate, and find out the feature subset which locate on the critical frequency bands for recognizing rat's vigilance states.**

***Index Terms*—feature selection, frequency band, pattern recognition, vigilance states**.

I. Introduction

Sleep is a physiological state composed of different stages. Electroencephalogram (EEG) analysis shows that typical patterns of activity are correlated with different stages of sleep, wakefulness, and some pathophysiological processes, such as seizures. For most researchers, it is important to identify sleep stages in some cases, for example, in sleep deprivation and seizure studies [1]-[5]. Typically, sleep stages can be identified by the combination of EEG, electromyogram (EMG), electroosculogram (EOG) and visual behavioral monitoring. However, it is a time-consuming task to score these vigilance states manually even when the analyzer is an expert.

In our pervious study [6], we proposed a machine learning method to classify there vigilance stages of rats. The vigilance stages are mainly categorized states into the following three states: the awake (AW) state, slow wave sleep (SWS) state, and rapid eye movement sleep (REM) state. During the AW state, the animal exhibits low amplitude and high frequency EEG results. Some investigators distinguish active awake from quite awake by high EMG activity. Figure 1 shows the EEG and EMG signals of the rat in our experiment.

Manuscript received November 06, 2012. This work was supported by the National Science Council, Taiwan, R.O.C., under the Grant NSC 101-2221-E-032-055.

Chien-Hsing Chou is with the Department of Electrical Engineering, Tamkang University, Taiwan (e-mail: chchou@mail.tku.edu.tw).
Chung-Chih Kuo is with the Institute of Neuroscience, Tzu Chi University, Taiwan.
Zong-En Yu is with the Department of Electrical Engineering, National Taiwan University, Taiwan.
Hsien-Pang Tai is with the Department of Electrical Engineering, Tamkang University, Taiwan.
Ke-Wei Chen is with the Department of Electrical Engineering, Tamkang University, Taiwan.

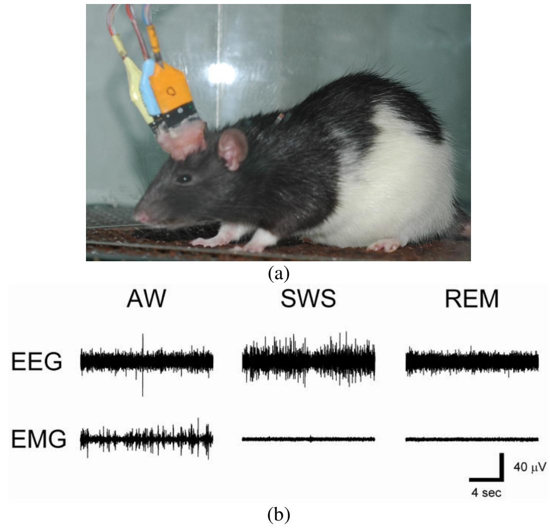

Fig. 1. (a) The rat in our experiment. (b) The EEG and EMG signals of three vigilance states.

In our pervious study [6], the proposed method could automatically recognize three vigilance states with high accuracy rate. However, the biological researchers more concern what are the meaningful frequency bands for classifying three vigilance states. To extract the meaningful frequency bands, feature selection algorithm is the useful technique to extract the critical feature subset (frequency bands) from the dataset. A successful feature selection not only improves classification accuracy but also extracts the critical features that users are concerned with. For instance, in the analysis research of DNA sequence, feature selection makes it possible to locate the segments on the sequence or the types of amino acids that may lead to certain diseases [7]-[8], or select the genes that may lead to certain diseases from the data of microarray [9]. Another example is text categorization, in which feature selection makes it possible to extract the keywords contributing to text classification [10]-[11]. In addition, feature selection only selects the features that users are most interested in, but also save time in training and testing the classifier, and reduce the memory space required for data storage.

To extract the critical frequency bands, we first transfer the EEG signal into frequency information with fast Fourier transform (FFT) [7]-[8]. The EEG spectrum is generated by using the FFT method with a 4-second window size, a frequency range from 0 to 50 Hz (see Fig. 2) Then the EEG spectrum is uniformly divided into 32 non-overlap frequency bands (see Fig. 3). The power of each frequency band is normalized by the sum of power for each frequency band. As a result, the EEG in each epoch can be transformed into 32 numerical features. We could then apply feature selection algorithm to extract the feature subset for classifying vigilance states. By examining the selected features of the feature subset, we can denote the critical frequency bands from these selected features.

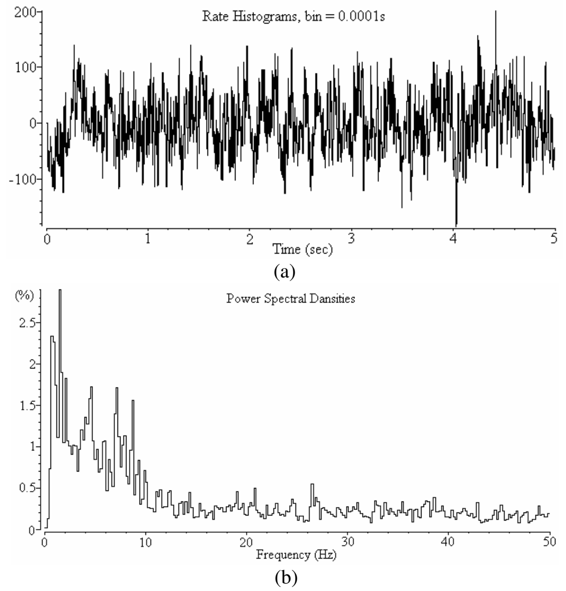

Fig. 2. (a) The EEG signal. (b) The EEG spectrum obtained by applying FFT.

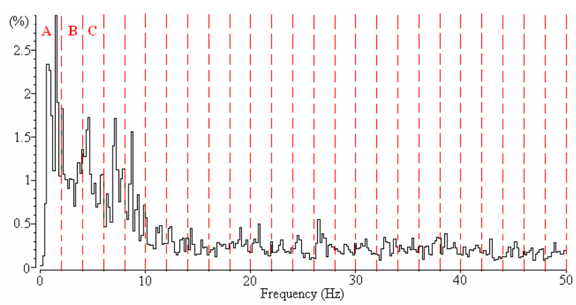

Fig. 3. The EEG spectrum is divided into a resolution of 32 frequency bands.

The proposed algorithm adopts the concept of neighborhood relation during adding and eliminating a candidate feature. The paper is organized as follows. In Section 2, the proposed feature selection algorithm is presented. In section 3, we introduce how to collect rat's EEG signals and generate the experimental dataset. In Section 4 the experimental results are used to demonstrate the effectiveness of the proposed algorithm. Section 5 concludes the paper.

II. The Proposed Neighborhood-Relation Feature Selection

In this section, we introduce the proposed neighborhood-relation feature selection algorithm (NRFS). The NRFS algorithm consists of two stages: add features and eliminate features. We apply sequential forward search algorithm (SFS) [9] to create an initial feature subset generated from an original feature set. Then, in the stage one of NRFS, we iteratively add candidate features to obtain better feature subset and higher recognition rate until if matches the stopping criterion. After adding features in stage one, the NRFS eliminates candidate features from the feature subset in stage two. To evaluate the recognition rate of the selected feature subset, the classification method used in this study is the k-nearest neighbor (kNN) [10]-[11] as the authentication method. The detailed steps of the NRFS algorithm are given as following.

Step 1: Initial Feature Subset by Using SFS

Use sequential forward search algorithm (SFS) to generate the initial feature subset for the NRFS algorithm.

Step 2: Calculate Weight Values of Unselected Features

To add the unselected features into the feature subset, we need calculate the weight values of these candidate features in advance. In this step, only calculate the weight value of unselected features (candidate feature for adding). Figure 4 is the diagram of the weight values, the red block denotes that this feature has been selected in the feature subset, and the white blocks are its neighboring features that are not selected. The unselected feature obtains a weight values (1 or 2) if it located at the neighborhood of the selected feature (red block). We use Figure 5 as the example to caculate the weight value of these unselected features.

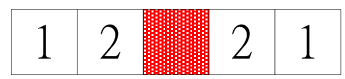

Fig. 4. The diagram of the weight values.

Feature Index	0	1	2	3	4	5	6	7	8	9
Weight Value	1	3			5		4		2	1

Fig. 5. The example to caculate the weight values of unselected features.

Step 3: Add Features

In this step, we sequentially add an unselected feature into the feature subset if the unselected feature has higher weight value. Feature '4' is first tried to add into the feature subset. If the recognition rate of the new feature subset could be improved or equal to original recognition rate; then the feature '4' is added into the feature subset; and go Step 2 to recalculate the weight values. Otherwise, try the next unselected feature according its ranking. If we can't further add any feature in this step, then go Step 4.

Step 4: Calculate Weight Values of Selected Features

Before eliminating the selected features from the feature subset, we need recalculate the weight values of these candidate features in advance. In this step, only calculate the weight values of selected features for eliminating in next step. The computing method applies the same diagram of the weight values as given in Fig. 4. We use Figure 6 as the example to caculate the weight values of these selected features. In this example, five features (feauter index '2', '3', '4', '5', and '7') have been selected in the feature subset. Each selected feature accumlates the weight value providing with its neighboring selected features. For example, the feature '3' obtains the highest weight value 5 by accumlating the weight values of three neighboring selected features (features '3', '4', and '5') , but the feature '7' only obtains the weight value 1 from the selected feature '5'.

Feature Index	0	1	2	3	4	5	6	7	8	9
Weight Value			3	5	5	4		1		

Fig. 6. The example to caculate the weight values of selected features.

Step 5: Eliminate Features

In this step, we sequentially eliminate a selected feature from the feature subset if it has lower weight value. As the example in Fig. 6, feature ‘7’ is first tried to eliminate from the feature subset. If the recognition rate of the new feature subset could be improved; then the feature ‘7’ is eliminated from the feature subset. If we can’t further eliminate any feature in this step, then end the NRFS algorithm.

III. Collect Rat’s EEG Signals and Generate the Experimental Dataset

In this section, we introduce how to collect the Rat’s EEG singles and generat the experimental dataset. The electrical activity of the EEG is a measure of the extracellular current flow from the activity of many neurons. This electrical activity originates from neurons in the underlying brain tissue. EEG oscillation depends on the activation timing of contributed neurons. In general, the frequency of normal EEG was distributed in the range of 1-50 Hz. The observed frequencies can be divided into several groups: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-20 Hz), and gamma (20-50) (Westbrook [12]). The major difference in spectrum patterns between the 3 states included the alpha, alpha and gamma band.

The spectrum of EEG in the AW state includes high alpha (8-13 Hz) and gamma (20-50 Hz) power levels [12]. The SWS state, which is defined by a high amplitude and low frequency EEG, begins with a sleep spindle and is dominated by delta (0.5-4 Hz) waves. In the REM state, the animal also shows similar characteristic of the AW state. High activity in the alpha and gamma bands are characteristics of REM state [12]-[14]. However, the animal is atonic and shows flat EMG activity.

Finally, 810 EEG epochs were collected and labeled as data patterns in the experimental dataset. The dataset is then partitioned into the training dataset and testing dataset. A total of 540 and 270 EEG epochs were used as the training and testing patterns, respectively. Table 1 lists the number of epochs in each state.

TABLE III: The number of epochs in each state.

State	Number of Training Epochs	Number of Testing Epochs
REM	56	41
SWS	248	155
AW	236	74
Total	540	270

IV. Experimental Result

The experimental simulations in this study compared the SFS and NRFS algorithms. Table 2 shows the simulation results. NRFS extracted more features and obtained a higher accuracy rate than SFS. Figure 7 show the feature subset selected by two algorithms. The SFS algorithm selected 6 features (Fig. 7(a)) into the feature subset. After apply NRFS algorithm, there are 10 features selected into the feature subset (Fig. 7(b)) and the accuracy rate is also improved. Besides, the feature selected by NRFS obviously located at two main regions. The frequency bands of the selected features represent the frequency for 11.2-14.4 Hz and 22.4-35.2 Hz. In which the frequency 11.2-14.4 Hz is identified as the alpha band, while the 22.4-35.2 Hz is identified as gamma band. The simulation result has the same conclusion to Westbrook’s research results.

TABLE II: Experimental results of Rat’s EEG signals.

	SFS	NRFS
Number of Selected Features	6	10
Accuracy of Testing Data (%)	92.65	94.03
Computational Time (sec.)	46	141

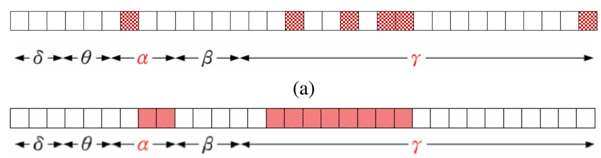

Fig. 7. Feature subset selected by (a) SFS; (b) NRFS.

V. Conclusion

In this study, we proposed the NRFS algorithm to extract the critical frequency bands for classifying vigilance states of rats. The proposed algorithm adopts the concept of neighborhood relation during adding and eliminating a candidate feature. The experimental result illustrates that the NRFS algorithm could find out the meaningful frequency bands the same with the results of several biological researchers.

References

[1] Balkin, T.J., O'Donnell, V.M., Kamimori, G.H., Redmond, D.P., & Belenky, G. , “Administration of triazolam prior to recovery sleep: effects on sleep architecture, subsequent alertness and performance.” *Psychopharmacology*, 99(4), pp. 526-531, 1989.

[2] Mendelson, W.B., & Bergmann, B.M. “Effects of pinealectomy on baseline sleep and response to sleep deprivation.” *Sleep*, 24(4), pp. 369-373, 2001.

[3] Bagshaw, A.P., Jacobs, J., LeVan, P., Dubeau, F., & Gotman, J. “Effect of sleep stage on interictal high-frequency oscillations recorded from depth macroelectrodes in patients with focal epilepsy.” *Epilepsia*, 50(4), pp. 617-628, 2008.

[4] Marzec M.L., Malow, B.A., “Approaches to staging sleep in polysomnographic studies with epileptic activity.” *Sleep Medicine*, 4(5), pp. 409-417, 2003.

[5] Natarajan, A., Marzec, M.L., Lin, X., Minecan, D., & Malow, B.A. “Interictal epileptiform discharges do not change before seizures during sleep.” *Epilepsia*, 43(1), pp. 46-51, 2002.

[6] Z. E. Yu, C. C. Kuo, C. H. Chou, C. T. Yen, F. Chang, “A Machine Learning Approach to Classify Vigilance States in Rats,” Expert Systems with Applications, vol. 38, no. 8, pp. 10153-10160, 2011.

[7] Cooley, J.W., & Tukey, J.W. “An algorithm for the machine calculation of complex Fourier series.” *Math. Comp.*, 19, pp. 297-301, 1965.

[8] Brigham, E.O., & Yuen, C.K. “The Fast Fourier Transform.” *IEEE Transactions on Systems, Man and Cybernetics*, pp. 146-146, 1978

[9] P. Pudil, J. Novovičová and J. Kittler, “Floating search methods in feature selection,” Pattern Recognition Letters, vol. 15, no. 11, pp. 1119-1125, 1994.

[10] A. Levine, L. Lustick, B. Saltzberg, “The nearest neighbor rule for small samples drawn from uniform distributions,” IEEE Trans. Information Theory vol. IT-19 no. 5, pp. 697-699, 1973.

[11] J. F. O'Callaghan, "An alternative definition for "neighborhood of a point," IEEE Trans. Computers, vol. C-24, no. 11, pp. 1121-1125, 1975.

[12] Westbrook, G.L. Seizure and Epilepsy. In E.R. Kandel, J.H. Schwartz, T.M. Jessell (Eds.) *Principles of Neural Science.* New York, 2000.

[13] Robert, C., Guilpin, C., & Limoge, A. "Automated sleep staging systems in rats." *Journal of Neuroscience Methods*, 88(2), pp. 111-122, 1999.

[14] Louis, R.P., Lee, J., & Stephenson, R. "Design and validation of a computer-based sleep-scoring algorithm." *Journal of Neuroscience Methods*, 133(1-2), pp. 71-80, 2004.

Chien-Hsing Chou received the B.S. and M.S. degrees from the Department of Electrical Engineering, Tamkang University, Taiwan, in 1997 and 1999, respectively, and the Ph.D. degree at the Department of Electrical Engineering from Tamkang University, Taiwan, in 2003. He is currently an assistant professor of electrical engineering at Tamkang University, Taiwan. His research interests include image analysis and recognition, mobile phone programming, machine learning, document analysis and recognition, and clustering analysis.

Chung-Chih Kuo is an assistant professor in Department of Physiology and Master Program of Anatomical and Physiological Medicine in School of Medicine at the Tzu Chi University. He earned a PhD in Zoology from the National Taiwan University in 2005. His research interests include sensory physiology, learning and memory, emotion, and drug addiction with electrophysiological technique. His work has been published in Pain, Journal of Neurophysiology, Brain Research, and Chinese Journal of Physiology. As a trained neuroscientist, he started his research on cortical representation of pain in conscious animals. His early work has focused on how the neurons of different brain areas responded to the noxious stimulation and compared the neuronal responses of brain areas in different components of pain. Dr. Kuo's recent work establishes an animal model to explore the mechanism of a pain therapy, motor cortex stimulation.

Zong-En Yu He is currently a PhD student at National Taiwan University, Taiwan. His research interests include machine learning, and neural science.

Hsien-Pang Tai received the B.S. and M.S. degrees from the Department of Electrical Engineering, Tamkang University, Taiwan, in 1997 and 1999, respectively. His research interests include image analysis and pattern recognition.

Ke-Wei Chen received the B.S. degree from the Department of Electrical Engineering, Tamkang University, Taiwan, 2011. He is currently a master student at Tamkang University, Taiwan. His research interests include image analysis and recognition and mobile phone programming.